The influence of degree of education in opinion about sexual relations with someone different than your marriage partner

Introduction:

This is a Data Analysis project for the Data Analysis and Statistical Inference course of Coursera. My main goal with this work, is to try to provide someevidence that **there is a relationship between the degree of education and the opinion about having sex with someone different than marriage partner**. I believe that studying the relationship between the degree of education and some behavioral or social aspects can be help to realize an approximation or distancing of some individual values, for example, in case of this study, betrayal. This work can be important for sociologists or psychologists, that want to have a better understanding of the human behavior associated with the education degree.

Data:

The data was collected via survey from 1972 until 2012. Until 1993, GSS was administered almost annually. The target sample size for the annual surveys was 1500; actual sample sizes ranged between 1372 (1990) and 1613 (1972). There were no GSSs in 1979,1981 and 1992. Some questions in the survey, are no applicable to some respondents. For example, the question on martial happiness is only asked of currently married respondents. For this study, each case represents the respondent of the survey. The variables analyzed by this study are:

- DEGREE - a categorical variable that represent the respondent's highest degree of education
- XMARSEX - a categorical variable that represent the opinion of respondents about having sexual relations with someone than the marriage partner

This is an observational study, because the data was collected by a survey and there is no interference in this data. The population of interests is a person resident in United States. Based on in the idea that Generalization is based on in the combination of two things: Random Sample and Random Assignment, the study can be generalized to the population of USA, but it's not Causal. There are some interesting source of bias in the study. First, the respondents can have the option of not answer a question, it means there is a Non-Response source of bias. Beside that, there is a kind of Convenience Sample, because the survey is mainly focused on household, this type of respondent is more accessible than others, for example, a person that needs go to the work every day. There is no causality in these variables, because there is no Random Assignment associated with the study.

Exploratory data analysis:

The objective of this study is trying to understand the relationship of two categorical variables: DEGREE and XMARSEX. Before starting to study more in depth this relationship, we will take a look on each column separately. Let's investigate the data in XMARSEX column:

```
table(gss$xmarsex, useNA = "ifany")
```

```
##
##     Always Wrong Almst Always Wrg  Sometimes Wrong Not Wrong At All
##            25929             4581             2652              857
##            Other             <NA>
##                0            23042
```

At the first impression, seems that the majority of the population thinks that the “having sexual relation with someone different than you marriage partner” is Always Wrong. Just a little part of this population think that is Not Wrong at All. The XMARSEX is categorical variable. To explore how representative are the answers for each option, we will build a frequency table for her. Take a look at the frequency table:

```
table(gss$xmarsex, useNA = "ifany")/dim(gss)[1]
```

```
##
##     Always Wrong Almst Always Wrg  Sometimes Wrong Not Wrong At All
##          0.45441          0.08028          0.04648          0.01502
##            Other             <NA>
##          0.00000          0.40381
```

As we said before, about 45% of the population, believes that is Always Wrong having sexual relation with someone different of his spouse and just 1.5% believes that is not Wrong at all. It's important to note that this variable have a significant number of NA's responses, about 40%, that can be justified by the fact that GSS survey was not executed in

some years. To verifiy that the NAs answers is related with the years that GSS survey was not executed, we can group the responses by years:

```
table(gss$xmarsex, gss$year, useNA = "ifany")
```

```
##
##                    1972 1973 1974 1975 1976 1977 1978 1980 1982 1983 1984
##   Always Wrong        0 1037 1082    0 1013 1103    0 1018 1333    0 1023
##   Almst Always Wrg    0  220  173    0  230  206    0  230  237    0  264
##   Sometimes Wrong     0  173  169    0  169  153    0  143  194    0  129
##   Not Wrong At All    0   61   36    0   63   48    0   53   57    0   33
##   Other               0    0    0    0    0    0    0    0    0    0    0
##   <NA>             1613   13   24 1490   24   20 1532   24   39 1599   24
##
##                    1985 1986 1987 1988 1989 1990 1991 1993 1994 1996 1998
##   Always Wrong     1132    0 1308  764  797  709  738  815 1546 1471 1466
##   Almst Always Wrg  207    0  263  125  131  115  132  152  248  285  231
##   Sometimes Wrong   130    0  165   54   74   63   63   60  130   98  107
##   Not Wrong At All   43    0   48   20   17   13   31   26   46   35   44
##   Other               0    0    0    0    0    0    0    0    0    0    0
##   <NA>               22 1470   35  518  518  472  553  553 1022 1015  984
##
##                    2000 2002 2004 2006 2008 2010 2012
##   Always Wrong     1448  725  717 1582 1088  988 1026
##   Almst Always Wrg  198  124  107  236  144  168  155
##   Sometimes Wrong   130   39   44  121   76   81   87
##   Not Wrong At All   47   19   19   36   20   26   16
##   Other               0    0    0    0    0    0    0
##   <NA>              994 1858 1925 2535  695  781  690
```

There were no GSSs in 1979,1981 and 1992. As we can see in the table, in these years we don't have information about the NAs, because the data were not inserted in this dataset. In this case, the higher number of NAs is not related with the years that GSS was not executed. A important information is that some questions may not exist in all years. Seems that the high number of NA is due to the fact that a question may not have been taken in that year. For example in 1972, 1978, 1983 and 1986 the NAs answer is higher, probably this question was not taken in these years. This is an important factor for high number of NA answers.

Now, lets do the same analysis to DEGREE variable. First, here a simple table of DEGREE variable data:

```
table(gss$degree, useNA = "ifany")
```

```
##
## Lt High School    High School Junior College       Bachelor       Graduate
##          11822          29287           3070           8002           3870
##           <NA>
##           1010
```

At the first impression, we can note that the degree of education of majority of the population is concentrated between Lt High School and High School Junior. Another important thing to note is the number of NAs responses, 1010. This number is much smaller than the XMARSEX, but it's still loud. Take a look at the frequency table for DEGREE variable.

```
table(gss$degree, useNA = "ifany")/dim(gss)[1]
```

```
##
## Lt High School    High School Junior College       Bachelor       Graduate
##        0.20718        0.51326        0.05380        0.14024        0.06782
##           <NA>
##        0.01770
```

Looking the data, we can see that the NA values represents about 1.7% of the total. Is much smaller than the 40% of the NAs in the XMARSEX variable. Now lets to analyze the DEGREE answers by year:

```
table(gss$degree, gss$year, useNA = "ifany")
```

```
##
##                  1972 1973 1974 1975 1976 1977 1978 1980 1982 1983 1984
##   Lt High School  587  510  479  492  482  488  431  407  539  385  374
##   High School     762  720  722  732  728  741  804  745  938  835  764
##   Junior College   17   20   25   33   24   34   42   45   74   61   54
##   Bachelor        124  132  139  140  152  139  151  158  165  202  175
##   Graduate         52   65   73   50   61   74   62   71   84   80   77
##   <NA>             71   57   46   43   52   54   42   42   60   36   29
##
##                  1985 1986 1987 1988 1989 1990 1991 1993 1994 1996 1998
##   Lt High School  387  376  441  335  315  268  280  267  478  425  409
##   High School     795  750  934  785  814  726  825  840 1586 1567 1500
##   Junior College   60   48   77   65   96   75   73   97  176  187  209
##   Bachelor        175  192  238  179  189  197  224  253  497  471  478
##   Graduate         90   80   87   89   95   87   81  118  216  224  205
##   <NA>             27   24   42   28   28   19   34   31   39   30   31
##
##                  2000 2002 2004 2006 2008 2010 2012
##   Lt High School  419  386  347  659  283  293  280
##   High School    1501 1485 1435 2273 1003 1001  976
##   Junior College  206  202  224  377  173  145  151
##   Bachelor        435  443  507  763  355  375  354
##   Graduate        218  230  281  403  194  218  205
##   <NA>             38   19   18   35   15   12    8
```

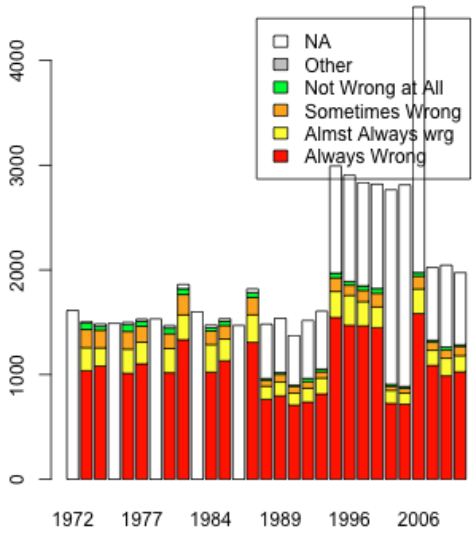
It seems that the DEGREE variable was answered in all years. It can explain why the NAs values are much smaller then in XMARSEX. Now lets take a look in the relationship of the these two variables:

```
table(gss$degree, gss$xmarsex)
```

```
##
##                  Always Wrong Almst Always Wrg Sometimes Wrong
##   Lt High School          5888              592             498
##   High School            13588             2224            1281
##   Junior College          1362              242             117
##   Bachelor                3191              920             427
##   Graduate                1369              563             289
##
##                  Not Wrong At All Other
##   Lt High School               208     0
##   High School                  398     0
##   Junior College                52     0
##   Bachelor                     103     0
##   Graduate                      85     0
```

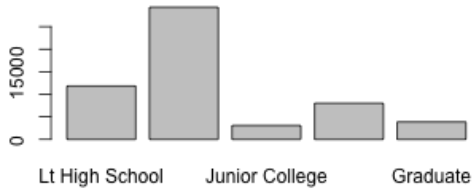
It seems that the majority of the population, independent of the degree, think that is Always Wrong having sexual relations with someone different than your marriage partner. Here, some visualizations to facilitate the understanding of the data: The distribution of XMARSEX answers by year:

```
barplot(table(gss$xmarsex, gss$year, useNA = "ifany"), col = c("red",
"yellow",
    "orange", "green", "grey", "white"), legend = c("Always Wrong", "Almst
Always wrg",
    "Sometimes Wrong", "Not Wrong at All", "Other", "NA"))
```

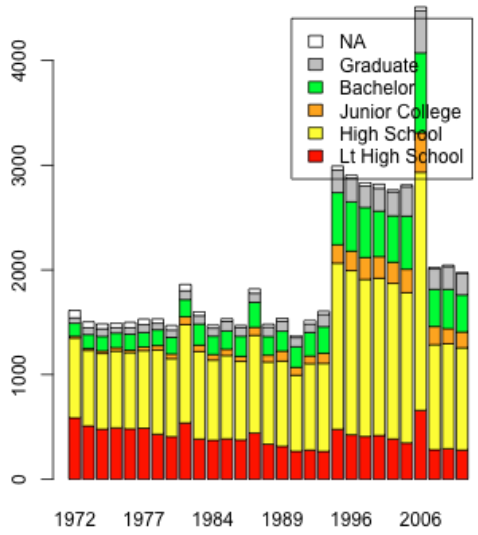


Some visualizations of DEGREE Variable:

```
plot(gss$degree)
```

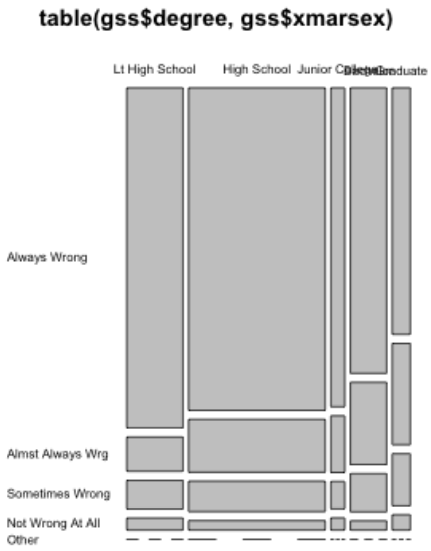



```
barplot(table(gss$degree, gss$year, useNA = "ifany"), col = c("red", "yellow",
    "orange", "green", "grey", "white"), legend = c("Lt High School", "High
School",
    "Junior College", "Bachelor", "Graduate", "NA"))
```



About the relationship between the two variables:

```
mosaicplot(table(gss$degree, gss$xmarsex), las = 1)
```



In the mosaic plot visualization, you can see that the distribution for the Always Wrong option is decreasing as the level of education increases. Despite having fewer responses, the Not Wrong At All option has greater number of responses to the highest degree of education.

Inference:

I want to check with my study if the degree of education influences the opinion about having sexual relations with someone different than your marriage partner. For verify this, I will use the **chi-square independent test**. We choose this method because we have 2 categorical variables with more then 2 levels. The first thing we need to do is build a contingency table for verify the relationship between the variables DEGREE and XMARSEX.

```
addmargins(table(gss$xmarsex, gss$degree))
```

```
##
##                     Lt High School High School Junior College Bachelor
##   Always Wrong                5888       13588            1362     3191
##   Almst Always Wrg             592        2224             242      920
##   Sometimes Wrong              498        1281             117      427
##   Not Wrong At All             208         398              52      103
##   Other                          0           0               0        0
##   Sum                         7186       17491            1773     4641
##
##                     Graduate   Sum
##   Always Wrong          1369 25398
##   Almst Always Wrg       563  4541
##   Sometimes Wrong        289  2612
##   Not Wrong At All        85   846
##   Other                    0     0
##   Sum                   2306 33397
```

The Hypotheses for this study are: **H0 (nothing going on)**: XMARSEX and DEGREE of education are **independent**. The opinion about having sexual relations with someone different than your marriage partner is not related with degree of education. **H1(something going on): XMARSEX and DEGREE of education are __dependent**. The opinion about having sexual relations with someone different than your marriage partner is related with degree of education. We need to check the conditions for the chi-squared test:

1) **Independence:**

- random sample/assignment ? Ok
- if sample without replacement, n < 10% of population? Ok
- each case only contributes to one cell in the table? Ok

2) **Sample size:** Each particular scenario (i.e cell) must have at least 5 expected cases? Ok

We do not execute this calculus by hand and we will use the Inference function in R to execute the chi-squared test for us. First, take a look again on the relationship of XMARSEX and DEGREE:

```
table(gss$xmarsex, gss$degree)
```

```
##
##                     Lt High School High School Junior College Bachelor
##   Always Wrong                5888       13588            1362     3191
##   Almst Always Wrg             592        2224             242      920
##   Sometimes Wrong              498        1281             117      427
##   Not Wrong At All             208         398              52      103
##   Other                          0           0               0        0
##
##                     Graduate
##   Always Wrong          1369
##   Almst Always Wrg       563
##   Sometimes Wrong        289
##   Not Wrong At All        85
##   Other                    0
```

Observe the row correspondent to the **Other** opnion for XMARSEX. There is no value for any column, just zero. To execute the chi-squared test correctly we need to remove all the 0 values for variables (see appendix).

Now, we are ready to use the inference function. First, we need to load the inference function:

```
load(url("http://s3.amazonaws.com/assets.datacamp.com/course/dasi/inference.Rdata"))
```

Now we will execute the function, for chi-squared independent test for a dropped NA's dataset:

```
inference(gss_with_degree_and_xmarsex_without_NA_columns$xmarsex,
gss_with_degree_and_xmarsex_without_NA_columns$degree,
    est = "proportion", type = "ht", method = "theoretical", alternative =
"greater",
    inf_lines = FALSE)
```

```
## Error: object 'gss_with_degree_and_xmarsex_without_NA_columns' not found
```

We have a p-value pretty small, p-value < 2.2e-16.

Conclusion:

We can conclude with this study, based on the chi-squared independence test and the value of p-value: **There is a relationship between the degrees of education and the opinion about having sexual relations with other than your marriage partner.** The p-value is pretty small, what means that we can reject the null hypothesis in favor of the alternative hypothesis.

References

- General Social Survey dataset, *Smith, Tom W., Michael Hout, and Peter V. Marsden. General Social Survey, 1972-2012 [Cumulative File]* . This is a sociological survey used to collect data on demographic characteristics and attitudes of residents of the United States. The dataset can be found online at: http://bit.ly/dasi_gss_data. The dataset cookbook can be found here: https://d396qusza40orc.cloudfront.net/statistics%2Fproject%2Fgss1.html#educ

Appendix

XMARSEX page:

```
head(gss$xmarsex, 200L)
```

```
##   [1] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [15] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [29] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [43] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [57] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [71] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [85] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##  [99] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [113] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [127] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [141] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [155] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [169] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [183] <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## [197] <NA> <NA> <NA> <NA>
## 5 Levels: Always Wrong Almst Always Wrg ... Other
```

```
tail(gss$xmarsex, 200L)
```

```
##   [1] <NA>              Always Wrong      <NA>              Always Wrong
##   [5] <NA>              Always Wrong      Always Wrong      Always Wrong
##   [9] Always Wrong      Sometimes Wrong   Always Wrong      <NA>
##  [13] Always Wrong      Always Wrong      Always Wrong      <NA>
##  [17] Sometimes Wrong   <NA>              Always Wrong      Always Wrong
##  [21] Always Wrong      Always Wrong      <NA>              Almst Always Wrg
##  [25] <NA>              Always Wrong      <NA>              Always Wrong
##  [29] Always Wrong      <NA>              Always Wrong      <NA>
##  [33] Almst Always Wrg  <NA>              Always Wrong      <NA>
##  [37] Always Wrong      <NA>              Always Wrong      Always Wrong
##  [41] <NA>              Always Wrong      Always Wrong      <NA>
##  [45] Always Wrong      Always Wrong      <NA>              Always Wrong
##  [49] Always Wrong      Always Wrong      Always Wrong      Always Wrong
##  [53] Always Wrong      Not Wrong At All  Always Wrong      <NA>
##  [57] Always Wrong      Always Wrong      <NA>              Always Wrong
##  [61] Always Wrong      Always Wrong      Always Wrong      Almst Always Wrg
##  [65] <NA>              Always Wrong      <NA>              Always Wrong
##  [69] <NA>              Always Wrong      Always Wrong      <NA>
##  [73] Always Wrong      Always Wrong      Almst Always Wrg  <NA>
##  [77] Always Wrong      <NA>              Sometimes Wrong   Always Wrong
##  [81] <NA>              Almst Always Wrg  Always Wrong      <NA>
##  [85] Always Wrong      <NA>              Always Wrong      <NA>
##  [89] <NA>              Always Wrong      Always Wrong      Always Wrong
##  [93] Always Wrong      <NA>              Always Wrong      Always Wrong
##  [97] <NA>              Always Wrong      <NA>              Always Wrong
## [101] <NA>              <NA>              Always Wrong      Always Wrong
## [105] Always Wrong      Always Wrong      <NA>              Always Wrong
## [109] <NA>              Always Wrong      Always Wrong      Always Wrong
## [113] <NA>              Always Wrong      Always Wrong      Always Wrong
## [117] Always Wrong      Always Wrong      Always Wrong      Always Wrong
## [121] Always Wrong      Always Wrong      Always Wrong      <NA>
## [125] Always Wrong      Always Wrong      <NA>              Always Wrong
## [129] Always Wrong      Always Wrong      Always Wrong      <NA>
## [133] <NA>              Always Wrong      Always Wrong      <NA>
## [137] Always Wrong      <NA>              Always Wrong      Always Wrong
## [141] <NA>              Always Wrong      Always Wrong      <NA>
## [145] Always Wrong      <NA>              Always Wrong      Always Wrong
## [149] <NA>              Always Wrong      Always Wrong      <NA>
## [153] Always Wrong      <NA>              <NA>              Sometimes Wrong
## [157] <NA>              Always Wrong      Always Wrong      <NA>
## [161] Always Wrong      Always Wrong      <NA>              Always Wrong
## [165] Always Wrong      <NA>              Always Wrong      <NA>
## [169] Sometimes Wrong   Always Wrong      <NA>              Always Wrong
## [173] <NA>              Always Wrong      <NA>              Always Wrong
## [177] <NA>              Always Wrong      <NA>              Always Wrong
## [181] <NA>              Always Wrong      <NA>              Almst Always Wrg
## [185] Almst Always Wrg  Almst Always Wrg  Always Wrong      <NA>
## [189] Almst Always Wrg  Always Wrong      <NA>              Sometimes Wrong
## [193] <NA>              <NA>              Always Wrong      <NA>
## [197] Always Wrong      Always Wrong      <NA>              Always Wrong
## 5 Levels: Always Wrong Almst Always Wrg ... Other
```

DEGREE page:

```
head(gss$degree, 200L)
```

```
##   [1] Bachelor       Lt High School High School    Bachelor
##   [5] High School    High School    High School    Bachelor
##   [9] High School    High School    High School    Lt High School
##  [13] Lt High School Lt High School Lt High School High School
##  [17] High School    Lt High School Bachelor       High School
##  [21] High School    High School    High School    High School
##  [25] Bachelor       High School    High School    High School
##  [29] High School    Lt High School Lt High School High School
##  [33] Bachelor       Lt High School High School    High School
##  [37] High School    Lt High School Lt High School High School
##  [41] Lt High School High School    Lt High School Lt High School
##  [45] Lt High School High School    High School    High School
##  [49] Lt High School Lt High School Lt High School High School
##  [53] Lt High School High School    Lt High School High School
##  [57] High School    High School    High School    Lt High School
##  [61] High School    Lt High School High School    Lt High School
##  [65] High School    Lt High School High School    High School
##  [69] High School    Bachelor       <NA>           High School
##  [73] Lt High School Lt High School High School    Bachelor
##  [77] Lt High School Lt High School High School    High School
##  [81] <NA>           High School    High School    High School
##  [85] High School    Bachelor       Graduate       Graduate
##  [89] Bachelor       High School    Bachelor       High School
##  [93] High School    High School    Bachelor       High School
##  [97] High School    High School    High School    High School
## [101] Lt High School High School    High School    High School
## [105] High School    Lt High School High School    High School
## [109] Bachelor       High School    Lt High School High School
## [113] Lt High School Lt High School High School    High School
## [117] Lt High School Lt High School Lt High School High School
## [121] Lt High School Lt High School Lt High School Lt High School
## [125] Lt High School Lt High School Bachelor       High School
## [129] Lt High School High School    Lt High School High School
## [133] High School    Lt High School Lt High School Lt High School
## [137] High School    Lt High School Lt High School High School
## [141] High School    High School    High School    Lt High School
## [145] Lt High School Bachelor       High School    Lt High School
## [149] Lt High School High School    High School    High School
## [153] Lt High School Lt High School High School    High School
## [157] Lt High School High School    High School    Lt High School
## [161] High School    Lt High School Lt High School Lt High School
## [165] Graduate       Bachelor       Bachelor       High School
## [169] High School    High School    Lt High School High School
## [173] Lt High School Lt High School Graduate       Bachelor
## [177] High School    Lt High School Bachelor       High School
## [181] Lt High School High School    Lt High School Lt High School
## [185] Lt High School Lt High School Lt High School High School
## [189] Bachelor       Graduate       High School    Graduate
## [193] Bachelor       High School    Lt High School High School
## [197] Lt High School High School    High School    High School
## Levels: Lt High School High School Junior College Bachelor Graduate
```

```
tail(gss$degree, 200L)
```

```
##   [1] High School    High School    High School    High School
##   [5] High School    High School    High School    High School
##   [9] Bachelor       High School    High School    Bachelor
##  [13] High School    Bachelor       High School    High School
##  [17] Bachelor       Lt High School Bachelor       Bachelor
##  [21] High School    Lt High School High School    High School
##  [25] Graduate       Bachelor       High School    High School
##  [29] Bachelor       High School    Lt High School High School
##  [33] Junior College Lt High School High School    High School
##  [37] Lt High School High School    High School    High School
##  [41] High School    Lt High School High School    Lt High School
##  [45] High School    High School    High School    High School
##  [49] High School    High School    Bachelor       High School
##  [53] High School    Graduate       High School    High School
##  [57] High School    Bachelor       Graduate       High School
##  [61] High School    High School    Junior College High School
##  [65] High School    High School    Bachelor       Bachelor
##  [69] High School    High School    Graduate       High School
##  [73] Bachelor       High School    High School    Bachelor
##  [77] High School    Junior College High School    Bachelor
##  [81] High School    High School    High School    Bachelor
##  [85] Graduate       High School    Junior College High School
##  [89] Lt High School Lt High School High School    Graduate
##  [93] High School    Bachelor       Lt High School Junior College
##  [97] Junior College <NA>           High School    High School
## [101] Junior College High School    High School    Graduate
## [105] High School    High School    Bachelor       Lt High School
## [109] High School    Lt High School Bachelor       Lt High School
## [113] Graduate       Lt High School High School    High School
## [117] High School    High School    High School    High School
## [121] High School    Lt High School Junior College High School
## [125] High School    Lt High School High School    High School
## [129] Lt High School Lt High School High School    High School
## [133] Bachelor       High School    High School    Lt High School
## [137] High School    High School    Lt High School High School
## [141] Lt High School Lt High School Lt High School Lt High School
## [145] High School    High School    High School    High School
## [149] Bachelor       Lt High School High School    Junior College
## [153] High School    High School    Lt High School High School
## [157] Lt High School <NA>           Graduate       High School
## [161] Junior College High School    Graduate       Bachelor
## [165] Lt High School Junior College High School    High School
## [169] High School    High School    High School    High School
## [173] High School    High School    High School    Bachelor
## [177] Bachelor       Bachelor       Junior College Junior College
## [181] Junior College High School    Bachelor       Graduate
## [185] Graduate       Bachelor       High School    High School
## [189] High School    High School    Graduate       Lt High School
## [193] Lt High School Lt High School Lt High School Bachelor
## [197] High School    High School    High School    High School
## Levels: Lt High School High School Junior College Bachelor Graduate
```

Steps to remove all the 0 values for variables

```
# just columns of interest, gss$degree and gss$xmarsex
gss_with_degree_and_xmarsex <- gss[c(12, 94)]

# remove rows with NAs in either column
gss_with_degree_and_xmarsex_without_NA_columns <-
gss_with_degree_and_xmarsex[!is.na(gss_with_degree_and_xmarsex[,
    1]) & !is.na(gss_with_degree_and_xmarsex[, 2]), ]

# droplevels
gss_with_degree_and_xmarsex_without_NA_columns <-
droplevels(gss_with_degree_and_xmarsex_without_NA_columns)

# observe that there is no NA values anymore
table(gss_with_degree_and_xmarsex_without_NA_columns$xmarsex,
gss_with_degree_and_xmarsex_without_NA_columns$degree)
```

```
##
##                     Lt High School High School Junior College Bachelor
##   Always Wrong                5888       13588            1362     3191
##   Almst Always Wrg             592        2224             242      920
##   Sometimes Wrong              498        1281             117      427
##   Not Wrong At All             208         398              52      103
##
##                     Graduate
##   Always Wrong          1369
##   Almst Always Wrg       563
##   Sometimes Wrong        289
##   Not Wrong At All        85
```